

ABSTRACT

Microarray is a device that can be employed in measuring o expression levels of thousands of genes simultaneously. It finally produces microarray data that contain useful information of genomic, diagnostic, and prognostic for researchers (Knudsen, 2002). Thus, there is a need to select informative genes that contribute to a cancerous state (Mohamad et al., 2009). However, the gene selection process poses a major challenge because of the following characteristics of microarray data: the huge number of genes compared to the small number of samples (higherdimensional data), irrelevant genes, and noisy data. To overcome this challenge, a gene selection method is used to select a subset of genes that increases the classifier's ability to classify samples more accurately (Mohamad et al., 2007). Recently, several methods based on particle swarm optimization (PSO) are proposed to select informative genes from microarray data (Chuang et al., 2008; Li et al., 2008; Shen et al., 2008). PSO is a new evolutionary technique proposed by Kennedy and Eberhart (Kennedy and Eberhart, 1995)]. It is motivated from the simulation of social behaviour of organisms such as bird flocking and fish schooling. Shen et al. (Shen et al. 2008) have proposed a hybrid of PSO and tabu search approaches for gene selection. However, the results obtained by using the hybrid method are less significant because the application of tabu approaches in PSO is unable to search a near-optimal solution in search spaces. Next, an improved binary PSO have been proposed by Chuang et al. (Chuang et al., 2008). This approach produced 100% classification accuracy in many data sets, but it used a higher number of selected genes to achieve the higher accuracy. It uses the higher number because of all global best particles are reset to the same position when their fitness values do not change after three consecutive iterations. Li et al. (Li et al., 2008) have introduced a hybrid of PSO and GA for the same purpose. Unfortunately, the accuracy result is still not high and many genes are selected for cancer classification since there is no direct probability relation between genetic algorithms (GA) and PSO. Generally, the proposed methods that based on PSO (Chuang et al., 2008; Li et al., 2008; Shen et al., 2008) are intractable to efficiently produce a nearoptimal (smaller) subset of informative genes for higher classification accuracy. This is mainly because the total number of genes in microarray data is too large (higher-dimensional data).